Wide Area Information Servers (WAIS): Concepts and Applications

Wide Area Information Servers (WAIS): Concepts and Applications
Script

1. Introduction

In a sense, people live by processing information. We share experiences, make decisions, and pass on knowledge by gathering and disseminating information in forms such as speech, books, signs, music, and images. Increasingly in modern society, information is represented in digital, electronic media that is far more flexible than traditional forms. Suddenly, we see the opportunity to take into account absolutely vast amounts of information and to analyze the information in ways that are extraordinarily complex. Our hope is that out of more complete information and more thorough analysis we will gain increased knowledge and better decisions.

1.0.1. Station ID

This presentation is a production of the United States Geological Survey, a bureau within the U.S. Department of the Interior. My name is Eliot Christian and I work for the U.S. Geological Survey in Reston, Virginia.

1.1. Purpose and Outline 1.1.1. Purpose

The purpose of this presentation is to introduce basic concepts associated with a new approach to information search and retrieval, known as "Wide Area Information Servers," or WAIS. It is intended for people with a wide range of computer skills. The concepts are presented first in a non-technical context, but some technical details are presented in addition for those who are interested. No one is expected to be familiar with all of the technologies discussed here, and you do not need to follow every discussion to understand the thrust of the ideas being discussed. When the technicalities get too thick, relax and wait a moment or two--the next discussion should be clearer.

1.1.2. outline

This presentation is in four sections, each of which has four parts within it. The four sections are titled: "Introduction", "Search and Retrieval", "Organizing Principles", and "Making It Happen". At the beginning of each section will be an outline of the four parts within that section. After the last section, there will be a quick review of the entire presentation.

The four parts of the Introduction section are:

(1) Purpose and Outline
(2) Information and Computers
(3) WAIS and Standards
(4) The Earth Science Data Directory

1.2. Information and Computers
1.2.1. information and data processing systems

Computers have proven to be excellent tools for helping people to organize and analyze data and information, and they have been used extensively in a variety of distinct fields. The library and information services community developed powerful techniques for handling textual information, such as bibliographic systems. At the same time, the data processing community developed powerful techniques for handling databases.

1.2.2. networks

On a separate track, communications systems have been created and deployed worldwide that allows these data and information systems to be accessed by millions of users. However, since each data and information system was developed separately, the specific system designs often have little in common in terms of user interactions. As people came to need access to many different data and information systems, the differences among systems often result in a frustrating experience for the user.

1.2.3. standards

The many thousands of sources for data and information worldwide represent a vast treasure waiting to be used. Now that users with personal computers have the means to access those systems over interconnected networks, the lack of commonality among the systems is the major barrier to the wealth of accumulated knowledge. Clearly, powerful international standards for information search and retrieval are needed desperately.

1.3. WAIS

1.3.1. history

Dow Jones News Service is one information services company that has a significant stake in simplifying access to information across many different sources. In 1990, the company asked Thinking Machines, Incorporated, to work on the problem. Since the goal was to simplify system interactions from the user's perspective, Apple Computers joined in to bring their expertise in user-friendly personal computing and the accounting firm of Peat Marwick provided the real world users to test the concepts.
Brewster Kahle, a co-founder of Thinking Machines, led the project that came to be known as "Wide Area Information Servers":
(Brewster speaks to how WAIS started, his realization that non-proprietary standards were needed, his going to library science school and approaching NISO to adopt and adapt Z39.50)

1.3.2. one question, multiple sources

In the WAIS approach, a user can ask questions in a single, consistent manner regardless of the specific information source. There are already hundreds of information sources with contents ranging over cooking, religion, poetry, weather, biology, earth science, and many other areas.

1.3.3. users in 18 countries

WAIS information sources exist in eighteen different countries world wide and there are tens of thousands of WAIS users.

1.3.4. NISO Z39.50

The success of WAIS depends critically on wide acceptance of the NISO 239.50 standard for information search and retrieval. WAIS developers actively participated in rapidly evolving the Z39.50 standard and the 1992 version of 239.50 is considered to be very powerful and forward-looking. While Z39.50 is a U.S. standard, there are international standards that are fully compatible with Z39.50.

The Z39.50 standard focuses on the interaction between two computers--one serving the user and the other handling an information source. Z39.50 standardizes how those computers interact and specifies the way in which a search and retrieval session proceeds. The standard does not specify how the source manages information nor how the user views information.

1.3.5. search request and response

After initialization is accomplished between the two computers, the Z39.50 standard specifies that a "search request" can be communicated in certain forms and that certain messages can be communicated in response to that search request. In a typical text search, the search request sends words to be found and the "search response" sends back the titles of documents that contain those words.

1.3.6. presentation request and response

The Z39.50 standard also specifies how the computers interact in retrieving information. When a user selects one of the documents to view, a "presentation request" is communicated. The actual contents of the document is returned as a "presentation response." As we will see later, the term "document" is a bit misleading since any digital object can be registered to Z39.50.

1.3.7. Z39.50 Implementors Group

In addition to WAIS itself, there are many information service providers actively involved in implementing the Z39.50 standard. Among the Z39.50 implementors are:

- AT&T
- Carnegie-Mellon University
- Chemical Abstracts Service
- Columbia University
- Florida Center for Library Automation

- Library of Congress
- Mead Data Central
- Massachusetts Institute of Technology
- Microelectronics Consortium of North Carolina
- Next Computer, Inc.
- On-line Computer Library Corporation
- Thinking Machines, Inc.
- University of California
- U.S. Geological Survey

1.4. Earth Science Data Directory
1.4.1. contents

The U.S. Geological Survey became involved with Z39.50 and WAIS through efforts to enhance the usefulness and availability of the Earth Science Data Directory (ESDD). ESDD is a compilation of descriptions of data holdings related to earth science. In addition to describing the data of the U.S. Geological Survey, there are descriptions of earth science data from all 50 states and a comprehensive set of descriptions related to the Arctic environment.

1.4.2. one of many

ESDD is one of hundreds of different data directories in existence worldwide, many of which would be of interest to users of the ESDD. Rather than attempt to create 'directories of directories,' we looked for an approach that would allow many directories to co-exist and be readily accessible without creating a logistical nightmare. Z39.50 is such an approach and the public domain WAIS software has been an excellent starting point.

1.4.3. Enhancements within NISO Z39.50

Tim Gauslin of the U.S. Geological Survey has adapted the public domain WAIS software to meet the needs of the ESDD:
(Tim talks about porting software to Data General, IBM PC DOS Windows, and mainframe. Also, phrase search, keyword search, and location search. Tim makes the point that all ESDD WAIS enhancements introduced are within the Z39.50 standard.)

This completes section 1: Introduction

2. Search and Retrieval

The second section of the presentation is "Search and Retrieval." The four parts of the Search and Retrieval section are:

(1) Full-text Search
(2) Browsing Sources
(3) Telecommunications
(4) Access to Networks

2.1. Full-text Search
2.1.1. Search Request

WAIS starts with a 'Query' screen.
The screens I'll be showing are from the MS-DOS/Windows version of the WAIS client software, developed by Tim Gauslin at the USGS. The particular design you'll see here is similar to the basic Macintosh and Unix versions of the client software screens.)
In the box labeled 'TELL ME ABOUT,' we type in our question and then hit the search button. We do not need to know 'valid keywords' or 'search-able fields' or 'Boolean operators' or 'database joins' or any other query language specifics. We just type in English and start the search.

2.1.2. Search response

In this example, we type in 'toxic metals' and the response pops up in the bottom part of the window. The WAIS response to a search request consists of a list of 'document titles' together with the size of each referenced document and its ranked score based on 'relevance.' In this case, the server generated relevance scores by using word occurrence only, but other schemes could be implemented as appropriate.

2.1.3. Presentation Request

At this point, we might want to view a document, so we would 'double-click' on a title to generate a WAIS 'presentation request.'

2.1.4. Presentation Response

A new window opens as the WAIS presentation response is returned. Since the document is of the 'text' type, WAIS opens a text display window and highlights the occurrences of my search words.

2.1.5. Show map

The USGS has added another feature to WAIS at this point. If the document text includes the words 'latitude' and 'longitude,' WAIS will place into the menu line a button labeled 'SHOW MAP.' Clicking on the 'SHOW MAP' button brings up an orthographic map centered on the area referenced in the text, with a polygon drawn between the latitude, longitude points. This data set concerns California, so we see a box around it on the map.

2.1.6. Relevance Feedback

Another powerful feature of WAIS uses the concept of 'relevance feedback.' Here, we've highlighted a part of the text that seems to be closer to our interests--say, toxic chemicals, emergency planning, Superfund, and so on. Rather than typing all of these words back into the 'TELL ME ABOUT' box, we simply do 'add section' and the highlighted section is included in our next query through the box label 'similar to' on the query window. We could also use the 'Add Doc' button to include an entire document as relevance feedback.

2.1.7. Document Scoring

When we do a search using relevance feedback, we can get back an overload of information. You can see here why we need to have a way of scoring the documents so that the ones most likely to be of interest percolate to the top of the list. In this response, the document ranked as most relevant is the one we used to generate the query.

2.1.8. Multiple Sources

In these examples, we've been using WAIS to access a single source--the Earth Science Data Directory maintained at the USGS. The real power of WAIS, though, is its ability to search across a wide range of information sources. To specify the sources to be searched by a WAIS query, we use this window and drag sources to the top box from the available sources in the bottom box. These 'available sources' are ones chosen as useful in our areas of interest.

2.2. Browsing Sources

Where do the available sources come from?
You find available sources using the same WAIS query we've been using to find entries in the Earth Science Data Directory. This is a key concept in WAIS--the sources are described by documents that are able to be searched just like any other document.

2.2.1. Directory of Servers

Although there isn't really a hierarchy among WAIS information sources, you do need to have a starting point for discovering sources. One such point is the 'Directory of Servers' maintained for WAIS users on the Internet.
Here's one example of a search against the 'Directory of Servers' to find sources having to do with 'earth science.'
You will discover that the range of subject matter among WAIS sources is extremely broad--including religion, cooking, and many others in addition to many dozens of science and technical information topics.

2.2.2. add sources

When you retrieve a document that has a 'source' type, you have the option to save it as an entry in your own list of sources available for searching. Saving a source entry records in your computer the addressing information needed to access the information sources.

2.3. Telecommunications

I'd like to spend a few moments discussing some basic concepts concerning how computer facilities are interconnected--topics within the field known as "Telecommunications."

2.3.1. Terminal/host connections

General purpose computers carry out instructions given them by people at some point. The computer user typically has a keyboard for input and a monitor screen to view the computer output. When the user's equipment is physically separate from the main computer and there are multiple users, we can call the main computer a "host" and the users equipment a "terminal." Connecting the terminal to the host is a communications facility such as a telephone line.

2.3.2. Client/server networks

No one computer can efficiently provide all the services you might want, so computers have become specialized. In a typical office setting, users each have personal computers plus there are computers devoted to managing large files and others devoted to managing expensive printers. Rather than have separate terminal/host connections between each possible pair of computers, all of the computers are often connected on a single line known as a network. Each computer providing services is known as a "server." From the server point of view, the user's computer is regarded as a "client", so we have a client/server relationship between computers on a network.

2.3.3. Protocols

Services are requested and delivered by computer-to-computer messages. The formal agreements about what the meaning of different messages is known as a "protocol". How messages are addressed and routed is the main distinction between different protocols. Unfortunately, most computers handle only one protocol and many of the frustrations in using computers today arise from protocol mismatches.

2.3.4. TCP/IP and OSI

Networks are often connected together, forming a "networks of networks". Worldwide, there are two dominant sets of inter-networking protocols--Open Systems Interconnection (OSI) and Transmission Control Protocol-Internetwork Protocol (TCP/IP). TCP/IP is the forerunner of OSI and is still used extensively among research and education institutions, especially on the network of networks known as the "Internet", and its emerging successor: the "National Research and Education Network" (NREN).

2.3.5. Network addressing

When messages among multiple computers are sent over a network, each message must include the address of both the sender and receiver. Let's look again at the "Sources Editor" in WAIS. For each WAIS server, we see an entry for "ip address". A unique address is assigned to each computer using TCP/IP on the Internet. Under the Setup option we can find the Internet address for the client computer--the one we are using to generate the search and retrieval requests.

2.3.6. Z39.50 and protocols

Although the examples in this presentation use WAIS sources on the Internet, the underlying Z39.50 standard can be used over OSI networks as well. In fact, the Z39.50 standard is at a higher level than any of the telecommunications protocols and so is compatible with all of them. Z39.50 can be used in a terminal/host relationship such as a dial-up telephone connection, or for access to local files such as those on a CD-ROM (digital data on a compact disk).

2.4. Access to Networks

Throughout the world, there are about one million computers connected to the Internet. Many of these computers also serve as hosts for dozens or hundreds of terminals--vastly expanding the number of people who can use Internet services such as electronic mail, file transfer, and remote computing.

2.4.1. Simple WAIS

WAIS implements Z39.50 as a client/server relationship using the TCP/IP protocol on the Internet. There are, however, WAIS clients that also act as host computers to allow access to WAIS from devices known as "dumb terminals." This access is called "Simple WAIS" since it supports only the simplest, text only, access to WAIS sources.

2.4.2. Bulletin Boards

Computers that host bulletin boards for dial-up users could extend their services by connecting to the Internet as WAIS clients and offering Simple WAIS to the customers. Those hosts could also become WAIS servers so that the contents of the bulletin boards would be available to anyone having WAIS access.

2.4.3. Other access

Often, potential WAIS users have a personal computer with a high-speed modem but do not have an Internet connection. These users could dial-up a WAIS client on another personal computer using software such as "Norton/PC-Anywhere" or "Novell's On-Net." Such software allows the user to employ WAIS as though he or she was actually sitting at the client computer. It is as though the user has "taken over" the WAIS client and so has all of the speed and connection capabilities of the client computer.

This completes section 2: Search and Retrieval

3. Organizing Principles

The third section of the presentation is "Organizing Principles." The four parts of the Organizing Principles section are:

(1) Digital Standards
(2) Beyond Text Retrieval
(3) Beyond Text Searching, and
(4) Organizing Information

3.1. Digital Standards

The usefulness of WAIS and Z39.50 ultimately depends on the value of the information that can be retrieved.

3.1.1. Type Registration

The Z39.50 standard requires that information providers support the delivery of text documents, but also provides a registration process for the delivery of more complex products (referred to in the standard as "document types"). In this way, availability of additional products does not compromise basic text search services.

3.1.2. Abstract Syntax Notation

Type registration requires that the digital product be described in a formal language called "Abstract Syntax Notation". This formal description assures compatibility across different types of computers. During the search and retrieval session, the client computer can determine when a server is capable of delivering additional digital products and can match those products to the client capabilities for handling such products.

3.1.3. Machine Readable Cataloging

One additional type already formally described and registered to Z39.50 is "BIB 1"-a description of records in the form of "Machine Readable Cataloging", or MARC. This type of record is internationally recognized for the exchange of bibliographic information among institutions such as libraries.

3.2. Beyond Text Retrieval

Certainly, there exists a huge store of textual information, but you would also like to tap into the vast array of other digital products, such as pictures, databases, computer programs, music, and video. Wherever these products are available in a form compatible with the client computer, servers could deliver the product directly.

3.2.1. Graphic Interchange Format (GIF)

Z39.50 Implementors each have a series of type codes reserved for testing prior to formal registration. Between WAIS clients and servers, a document type has been defined for pictures in the form known as "Graphic Interchange Format" (GIF).

3.2.2. Weather Source

The usefulness of the GIF type is seen in the Weather Source available to WAIS users. From this information source, you can retrieve text descriptions of the weather forecast for selected cities nationwide. You can also receive the weather satellite picture as you might see on the evening news, plus the analysis maps prepared by the National Weather Service.

3.2.3. Hypermedia

WAIS and other Z39.50 Implementors will be rapidly expanding the range of products that can be retrieved. Since software is itself a digital product, it will be possible to deliver not only static products but interactive systems of information such as hypermedia.

3.3. Beyond Text Searching

Many sources of information are able to be searched with a text-based search, but there are also cases where searching strictly by text is not very appropriate.

3.3.1. Location Searching

If you are interested in finding data and information pertaining to a particular place, you might begin by searching on a place name. Yet, depending on your needs and your sources, you may not know all of the names you ought to use to find everything relevant and in any case the search can be tedious. The USGS has worked with the University of North Carolina to add into WAIS the ability to search data for a location on earth by simply drawing on a map. Here we've drawn a figure around Alaska and we see the corresponding latitude/longitude pairs in the 'Tell me about' box.

3.3.2. Structured Query Language

Many sources are basically compilations of numeric data. In these cases, the data are often arranged to be retrieved by a facility known as "Structured Query Language" (SQL). SQL queries are expressed in text and can be readily handled in Z39.50. However, SQL presumes the user already has some information about the various field names and organization of the data tables. For a novice user trying to use SQL directly in Z39.50, it may be appropriate to provide documentation with a fill-in-the-blanks approach and have the server take queries as relevance feedback. Other approaches to handling SQL within Z39.50 are being actively pursued, as well.

3.3.3. Language Translation

The current text searching orientation of Z39.50 could be extended in a variety of ways. For example, in looking for matches on text patterns, it is assumed that searchers write in the language of the source. Yet, already we have many sources in French, Russian, and other languages. The handling of multiple languages will greatly improve access to information, not only across natural languages but across the language barriers of technical jargon.

3.3.4. Other Patterns

Searching for patterns in text is relatively straightforward, but it is only one narrow application of pattern searching. Eventually, we would like to highlight a part of a picture and ask the client to find more pictures like it. Entire archives such as the vast holdings of Landsat imagery or the accumulated video imagery within broadcast news organizations might become readily accessible. Already, the WAIS software is being used to provide quick and cheap access to scanned images of publications.

3.4. Organizing Information

Given that most of us are deluged with huge amounts of advertising and junk mail, it may seem strange to advocate getting vastly more information. The value of information doesn't grow in relation to its volume--the most valuable information is that which is perceived as succinct and relevant by the user. Since the user has the best chance of knowing what he or she wants, it makes sense that users should be involved in organizing and filtering the information. Rather than being passive recipients of information, users can take the role of active information seekers--provided they have computers to handle the tediousness of the task.

3.4.1. Organized but not centralized

At first glance, it may seem that there ought to a single, master directory out on the network where you could go to find who has the information you need. But, how would such a directory be constructed? Would those who built it really know exactly how you are going to want to search it? In fact, don't your own information needs change over time and across your many interests?

3.4.2. sources as server directories

We have already seen that there are directories of servers for WAIS sources. Some sources are described in more than one directory, and those descriptions might read quite differently depending on who is expected to reference the directory. For example, the earth science data directory may be listed in several directories: Federal government, state government, geography, hydrology, geology, and global change research. In effect, those who create directories to other sources are providing an endorsement in the expectation that searchers would be guided by their opinions. This is equivalent to the services offered by journal editors or book reviewers in print media. We can expect healthy competition among analysts providing directory services just as we have competition among analysts dealing in stock market information.

3.4.3. Other navigation tools

There are various software tools for finding information on the Internet--George Brett of the Microelectronics Consortium of North Carolina: *(George Brett describes the variety of Internet navigation tools such as Archie, Gopher, Prospero, World Wide Web, and makes the point that WAIS will be a unifying theme.)*

3.4.4. Security, authentication, and charging

The Z39.50 standard provides for authentication so that servers can challenge users to provide the appropriate password in case the information is sensitive or there is a charge for access. WAIS sources on the Internet are typically free, in keeping with the traditional research and academic orientation of the Internet. Users should not be surprised to eventually find some misinformation sources and some fairly trashy material mixed in with the quality services. The Internet is a public forum and efforts to raise the level of discourse often founder on a widespread aversion to censorship.

This completes section 3: Organizing Principles

4. Making It Happen

The fourth section of the presentation is "Making It Happen." The four parts of the Making It Happen section are:

(1) Source Ideas
(2) Directory Ideas
(3) Requirements, and
(4) Creating Sources

4.1. Source Ideas

There are many different applications of WAIS being pursued all over the world. The following are ideas being pursued.

4.1.1. Apple Rosebud

(_____ of Apple Computers describes the Rosebud and Reporter projects intended to create a "personalized newspaper" using WAIS.)

4.1.2. National Geographic Data System

The WAIS approach is being pursued to build a national infrastructure for mapping data--Gene Thorley of the U.S. Geological Survey: *(Gene Thorley describes the WAIS vision for the 'National Geographic Data System.')*

4.1.3. electronic mail/bulletin board

(Doug Nebert describes how WAIS is being used to provide access to frequently asked questions about the ARC/INFO Geographic Information System.)

4.1.4. Word Perfect

(Jim Stapleton describes using WAIS to access Word Perfect documents in office files.)

4.1.5. CD-ROM's

(Jerry McFaul describes how WAIS can work with CD-Rom's.)

4.1.6. Contributor's Tool Kit

(Tim Gauslin describes how the idea of using WAIS as part of a "Contributor's Tool Kit".)

4.1.7. imaging;

(Brewster describes use of WAIS for access to scanned images.)

4.1.8. USGS NWIS and DSDL

(Owen Williams describes WAIS and the National Water Infromation System.)

4.2. Directory Ideas

The concept of directories is being extended using WAIS, as the following descriptions illustrate:

4.2.1. HPCCI Software Directory

(Barry Jacobs on the Software Directory being built as part of the U.S. High Performance Computing and Communications Initiative.)

4.2.2. EOSDIS

(Judy Feldman describes the role of WAIS in the future data and information system for the Earth Observing System, also known as "Mission to Planet Earth".)

4.2.3. NTIS FEDLINE

(Don Johnson describes WAIS and FEDLINE, a directory facility being established under law to provide for public access to government data and information.)

4.2.4. GCDIS

(McClure describes WAIS and the proposed Government-wide Information Inventory/Locator System under study by the Office of Management and Budget.)

4.2.5. GPO

(_____ describes WAIS and the role of the Government Printing Office in making available to the public Federal publications and related information.)

4.2.6. UNEP/IGBP/Agenda 21

(Hassan Virji describes WAIS and its potential utility in both the International Geosphere/Biosphere Programme and the United Nations Environment Programme.)

4.2.7. NASA NAM

(Gladys Cotter describes WAIS and NASA's new system to consolidate access to huge volumes of technical reports under the banner of "NASA Access Method," NAM.)

4.2.8. USGS DSDL

(Doug Nebert describes WAIS and the Distributed Spatial Data Library.)

4.2.9. EPA Envirofacts

(Al Pesachowitz describes WAIS and EPA Envirofacts.)

4.2.10. NARA AIS

(Ken Thibodeau describes WAIS and the new Archives Information System, AIS.)

4.3. Requirements

4.3.1. Clients

The public domain WAIS software is freely distributed to operate on a variety of computers. The software that a user would need (the "client" software) is available for Unix workstations in X-Windows. It is also available for MS-DOS computers (also knwon as IBM-compatible personal computers). For MS-DOS, separate versions are available depending on whether or not the user has MicroSoft Windows 3.0 or later. The software is also available for the Apple MacIntosh and it comes with the NeXT computer.

4.3.2. Servers

If you are interested in publishing a WAIS infromation source, you will need to use the public domain WAIS server software. It too is freely distributed and runs on Unix computers, plus the Digital Equipment Corporation VAX computers, IBM mainframe computers under the VM or MVS operating systems, and the massively parallel Connection Machine from Thinking Machines, Incorporated. The list of computers is probably out of date as you view this, since WAIS availability is expanding rapidly.

4.3.3. indexing time

Indexing of text to create an information source is fairly rapid--a 30 megabyte file was indexed in about twenty minutes on an inexpensive Unix workstation. A 30 megabyte file represents about 15,000 pages of typical typewritten text.

4.3.4. responsiveness

In the cases explored thus far, searching occurs in a matter of seconds, whether locally or over the Internet. The repsonsiveness of the server is not strongly affected by how much infromation it holds, but is affected by how many words are being searched at a time.

4.3.5. Commuications software

WAIS clients and servers have been implemented without communications capability on Unix, Macintosh, and MS-DOS. This allows WAIS to be used to access data and information distributed on floppy disks or CD-ROM's. Accessing WAIS directly on the Internet requires TCP/IP software. TCP/IP software is usually included with Unix workstations. Freely distributed, public domain TCP/IP software is avilable for the Macintosh, and for the MS-DOS versions of WAIS.

4.4. Creating Sources

4.4.1. Indexing software

The public domain WAIS package includes assistance for creating information sources. Indexing software is provided in the WAIS package for several document types (e.g., free text, Graphic Interface Format). The source code for the indexer is provided and is designed to be easily modified for adding other document types.

4.4.2. Interface routines

If access to databases other than those created by the indexer is required, the server interface routines are also designed to be customized. A typical customization would be to use search requests to construct an SQL query for a relational data base such as Ingress or Oracle.

5. Wrap Up

5.1. Review outline

5.1.1. Information

Computers have proven to be excellent tools for helping people to organize and analyze data and information. We now have the opportunity to take into account absolutely vast amounts of information and to analyze the information in ways that are extraordinarily complex. However, the lack of commonality among the systems is the major barrier to the wealth of accumulated knowledge. Clearly, powerful international standards for information search and retrieval are needed desperately.

5.1.2. WAIS and Standards

The WAIS approach is built on the NISO 239.50 standard for information search and retrieval. Z39.50 standardizes how computers interact--it does not specify how the source manages infromation nor how the user views information. There are dozens of major corporations and hundreds of universities actively implementing Z39.50 and/or WAIS. WAIS information sources exist in eighteen different countries world-wide and there are tens of thousands of WAIS users.

5.1.3. Data Directories

The U.S. Geological Survey's Earth Science Data Directory is a compilation of descriptions of data holdings related to earth science. It is one of hundreds of different data directories in existence world-wide. Rather than attempt to create 'directories of directories,' the USGS adopted an approach that allows many directories to co-exist and be readily accessible. Z39.50 is the standard that accommodates such an approach and the public domain WAIS software is an excellent way to implement Z39.50 quickly and inexpensively.

5.1.4. Searching and retrieving

In WAIS, we select the sources to be searched and then just type in the question in English and start the search

The USGS has added a location searching feature to WAIS that allows you to ask for data about an area on a map. Another powerful feature of WAIS uses the concept of 'relevance feedback.' Rather than typing all of the words that define what you're interested in, you simply highlight pieces of documents as you read them. Since WAIS typically returns many documents per search, ranking the documents by their relevance eases the task of scanning so much information.

5.1.5. Sources and Directories

Remember that WAIS sources are described by documents that are able to be searched just like any other document. Also, groupings of WAIS sources can be described as a source that we refer to as a directory of servers--one of which is the 'Directory of Servers' maintained for WAIS users on the Internet.

5.1.6. Access to Networks

Worldwide, there are two dominant sets of inter-networking protocols--Open Systems Interconnection (OSI) and Transmission Control Protocol-Internetwork Protocol (TCP/IP). Throughout the world, there are about one million computers connected to the Internet, which use TCP/IP. Although the examples in this presentation use WAIS sources on the Internet, the underlying Z39.50 standard can be used over OSI networks as well. Z39.50 can be used in a terminal/host relationship such as a dial-up telephone connection, or for access to local files such as those on a CD-ROM (digital data on a compact disk). Also, computers that host bulletin boards for dial-up users could extend their services by connecting to the Internet as WAIS clients and offering Simple WAIS to the customers. Those hosts could also become WAIS servers so that the contents of the bulletin boards would be available to anyone having WAIS access.

5.1.7. Beyond Text

The Z39.50 standard requires that information providers support the delivery of text documents, but also provides a registration process for the virtaully any kind of digital product, such as pictures, databases, computer programs, music, and video. One such type is an international standard format for the exchange of bibliographic information among institutions such as libraries.

5.1.8. Requirements

WAIS client software is freely distributed to operate on a variety of computers, including Unix workstations, MS-DOS personal computers, and the Apple MacIntosh. WAIS server software runs on Unix computers, VAX computers, IBM mainframe computers, and the Connection Machine, among others.

5.2. Further information

5.2.1. MCNC

(George Brett describes MCNC support.)

5.2.2. WAIS, Inc.

(Brewster Kahle to describe WAIS, Inc. products and services, relationship to the Z39.50 Implementors Group, and to the new MCNC. "Clearinghouse for Networked Information Discovery and Retrieval")

5.2.3. USGS

The U.S. Geological Survey will continue to support the implementation of Z39.50 for the Earth Science Data Directory and the National Geographic Data System.

Documentation on WAIS is available by "anonymous ftp" from "quake.think.com" or write to:

WAIS Documentation
U.S. Geological Survey
802 National Center
Reston, Virginia 22092

Thank you

Relevant Events:

Nov. 5	**WAIS Class at USGS (Library, NLM, EPA, NAL)**
Nov. 11	**Brewster Kahle in Reston**
Nov. 19	**Coalition for Networked Information at Lands Down (near Dulles), includes Brett, McClure, possibly Kapor**
Dec. 10	**Internet Executive Seminar downtown Washington includes WAIS demo, McClure, possibly Brett**